

Aston S, Hurlbert A. [What #theDress reveals about the role of illumination priors in color perception and color constancy](#). *Journal of Vision* 2017, 17(9).

**Copyright:**

© 2017 The Authors

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](#).

**DOI link to article:**

<http://doi.org/10.1167/17.9.4>

**Date deposited:**

07/11/2017



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

# What #theDress reveals about the role of illumination priors in color perception and color constancy

Stacey Aston

Institute of Neuroscience, Newcastle University,  
Newcastle upon Tyne, UK



Anya Hurlbert

Institute of Neuroscience, Newcastle University,  
Newcastle upon Tyne, UK



The disagreement between people who named #theDress (the Internet phenomenon of 2015) “blue and black” versus “white and gold” is thought to be caused by individual differences in color constancy. It is hypothesized that observers infer different incident illuminations, relying on illumination “priors” to overcome the ambiguity of the image. Different experiences may drive the formation of different illumination priors, and these may be indicated by differences in chronotype. We assess this hypothesis, asking whether matches to perceived illumination in the image and/or perceived dress colors relate to scores on the morningness-eveningness questionnaire (a measure of chronotype). We find moderate correlations between chronotype and illumination matches (morning types giving bluer illumination matches than evening types) and chronotype and dress body matches, but these are significant only at the 10% level. Further, although inferred illumination chromaticity in the image explains variation in the color matches to the dress (confirming the color constancy hypothesis), color constancy thresholds obtained using an established illumination discrimination task are not related to dress color perception. We also find achromatic settings depend on luminance, suggesting that subjective white point differences may explain the variation in dress color perception only if settings are made at individually tailored luminance levels. The results of such achromatic settings are inconsistent with their assumed correspondence to perceived illumination. Finally, our results suggest that perception and naming are disconnected, with observers reporting different color names for the dress photograph and their isolated color matches, the latter best capturing the variation in the matches.

## Introduction

The division between people who named #theDress (the dress photograph that first appeared on social

media in February 2015; Figure 1) “blue and black” versus “white and gold” illustrates the subjectivity and individuality of color perception. While there are many examples of illusions in which an individual observer sees a colored object differently in different viewing conditions (e.g., color contrast and color assimilation; Brainard & Hurlbert, 2015), #theDress phenomenon differs from these in eliciting striking inter-individual differences under identical viewing conditions.

When #theDress first appeared people immediately fell into two groups: one group reported a blue dress with black lace and the opposing group saw a white dress with gold lace. As controlled studies have since shown, these naming differences are not due to different viewing devices (Gegenfurtner, Bloj, & Toscani, 2015; Lafer-Sousa, Hermann, & Conway, 2015). In addition, the stark division of people into only two naming groups seems to have occurred only due to the way the question was posed on social media. When participants were allowed to free-name the colors of the dress, a continuum of color names emerged, but with three modal groups: blue and black (B/K), white and gold (W/G), and blue and gold (B/G; Lafer-Sousa et al., 2015).

Previous studies have also shown that observers differ not only in how they name the dress but also in the colors to which they match it (Chetverikov & Ivanchei, 2016; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015), leading to the conclusion that the phenomenon is a perceptual one and therefore not explained solely by differences in color naming and/or color categorization. However, previous studies differ somewhat in their matching results. In one study, where 53 laboratory participants made matches to the dress body and lace using a color picker tool, matches to both regions differed between B/K and W/G observers in both lightness and chromaticity (Lafer-Sousa et al., 2015). In another, Gegenfurtner et al.

Citation: Aston, S., & Hurlbert, A. (2017). What #theDress reveals about the role of illumination priors in color perception and color constancy. *Journal of Vision*, 17(9):4, 1–18, doi:10.1167/17.9.4.

doi: 10.1167/17.9.4

Received January 10, 2017; published August 9, 2017

ISSN 1534-7362 Copyright 2017 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/936403/> on 11/07/2017



Figure 1. The photo of the dress taken from Wikipedia. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

(2015) asked participants to make color matches to both the dress body and lace as well as to select the best matching chip from the glossy version of the Munsell Book of Colour. Contrary to the results of Lafer-Sousa et al. (2015), both sets of matching data did not differ in chromaticity between the two groups of observers but did differ in luminance (or value in the case of the Munsell chips). A later online survey found differences in the matches for the B/K and W/G observer groups along all axes of CIELAB color space (Chetverikov & Ivanchei, 2016). It is clear from these results that representing an individual's perception of the photograph only by the color names they assign is not enough to capture the variability in perception across the population. Clearly, there are substantial individual differences in the processes of color perception that are responsible for the phenomenon, but the detailed characteristics of these underlying factors remains somewhat elusive.

Brainard and Hurlbert (2015) suggested the color constancy hypothesis to account for the differences. Color constancy is the phenomenon by which the perceived colors of objects remain stable despite changes in the incident illumination (Hurlbert, 2007; for review see Foster, 2011, or Smithson, 2005). The perceptual mechanisms underlying color constancy are generally thought to involve an unconscious inference about the incident illumination, which is then discounted to enable the recovery of constant surface properties. The color constancy hypothesis for #theDress phenomenon implies that individual differences in perception of the photograph arise because observers make different inferences about the incident illumination spectrum in the scene. The information that the photograph itself provides about the illumination spectrum is ambiguous, because the colors in the photograph could be produced by illuminating a blue dress with yellow light or a white dress with blue

light (see figure 2 in Brainard & Hurlbert, 2015). If a difference in inferred illumination is responsible for the difference in perception then we must ask, what causes different observers to infer different illuminations? It might be that observers rely on previous experience—unconsciously embedded in the perceptual process—to determine the most likely illumination and overcome the uncertainty in the visual information.

In the Bayesian framework for visual perception, the knowledge learned from previous visual experience is incorporated as prior probability distributions on the stimulus space, or “priors.” Under the Bayesian brain hypothesis, it is assumed that the visual system uses priors to weight the uncertainty of sensory signals, effectively biasing perception towards the observer's internal expectations based on prior experience with stimuli (Allred, 2012; Brainard & Freeman, 1997; Kersten & Yuille, 2003; Sotiropoulos & Seriès, 2015). For example, internal priors may bias strawberry shapes to be red, or more relevant to the current study, daylight illuminations to vary from blue to yellow. Although there is evidence for the contribution of prior knowledge to color perception (Allred, 2012; Hansen, Olkkonen, Walter, & Gegenfurtner, 2006), there is a lack of evidence for illumination priors per se. Previous work has shown that the human visual system displays a robust color constancy bias for bluer daylight illumination changes, with observers finding it harder to discriminate an illumination change on the scene if the illumination becomes bluer rather than yellower, redder, or greener (Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Radonjic, Pearce et al., 2016). It is unresolved whether this reduced sensitivity to bluer illumination changes is a top-down influence on visual perception, as priors are generally considered to be, or whether the lack of sensitivity to these changes has become embedded in bottom-up visual processing through development and/or evolution. In either case, it is possible that individual differences in illumination estimation or illumination change discrimination may be due to differences in experience. If two individuals experience different occurrence distributions of illumination spectra, leading to the formation of different illumination priors, or to their visual systems adapting to become insensitive to these changes, the consequence in either case will be a difference in immediate perception.

If experience does play a role in how the visual system calibrates itself for environmental illumination changes, then observers with different experiences of illumination changes may exhibit different perceptions. To test this reasoning, one needs a measure that predicts behavioral exposure to particular illuminations. Given that daylight illumination chromaticities

are known to vary throughout the day (Hernández-Andrés, Romero, Nieves, & Lee, 2001; Wyszecki & Stiles, 1967) and traditional indoor lighting is much yellower than daylight (in particular for incandescent bulbs, see figure 2 in Webb, 2006), it is plausible that the observer's illumination prior may be conveyed by the observer's chronotype (Lafer-Sousa et al., 2015), where chronotype refers to whether an individual is a morning or evening type (more colloquially, whether an individual is a lark or an owl). Morning types might be more likely to experience the blues of daylight illuminations while evening types might be more likely to experience the yellows of artificial illuminations.

The color constancy explanation of the dress phenomenon is partially supported by the recent finding that the illumination inferred by observers in the photograph is negatively correlated with their dress color matches (yellower illumination implies bluer dress match and vice versa; Witzel, Racey, & O'Regan, 2017). However, Witzel et al. (2017) concluded that differences in inferred illumination chromaticity are not due to illumination priors, but rather because each observer implicitly estimates the illumination on a scene in an ad hoc manner. This interpretation implies that two observers who fall into different color-naming categories for the dress (e.g., blue/black vs. white/gold) will not display other distinct and predictable behavioral characteristics due to different inherent illumination priors. The results we present here allow for the possibility that illumination priors are responsible for differences in inferred illumination chromaticity, and hence bias dress color perception, given that we find a correlation between chronotype and color matching data. This correlation is weak and nonsignificant at the 5% level. On the other hand, we show that color constancy thresholds obtained using an established illumination discrimination task are not related to dress color perception, implying that generic biases in color constancy do not explain variations in dress color perception.

In addition, we show a dependence on luminance of subjective achromatic settings, which leads to the conclusion that subjective white point settings may be predictive of dress color perception in the photograph, but only if the settings are made at a luminance level that represents each individual's perceived level of brightness in the scene. Moreover, we present results that suggest perception and naming are disconnected, by showing that the color names observers report for the dress photograph differ from those that they give to their dress matches when shown them in isolation, with the latter best capturing the variation in the matches.

## Methods

### Overview

Participants completed a set of computer-based matching tasks in which they provided color appearance matches to the dress body and lace, matches to the perceived illumination in the image, and achromatic settings for an isolated disk. After completing the matching tasks, participants were shown their matches to the dress body and lace and asked to name them (disk color names). They were also asked to report (without restrictions) the colors that they named the dress the first time they saw it (dress color names). Observers who had not seen the dress before participating in the study were shown the dress photo and asked to name the colors of the dress. Each participant also completed the morningness-eveningness questionnaire (MEQ; Horne & Ostberg, 1976) as a measure of chronotype. In addition, all participants completed an illumination discrimination task to measure color constancy thresholds. The study was conducted between the months of January and July 2016. (Note that hereafter, the phrase “the dress” refers to the image of the dress in the original photograph).

### Participants

Participants were recruited from the Institute of Neuroscience (Newcastle University) volunteer pool, undergraduate courses, and by word of mouth. Thirty-five participants were recruited in total. Two participants were unable to complete the matching task without the aid of an experimenter who used the Xbox controller to adjust the patch according to the participant's instructions. Experimenter bias could not be ignored in these cases and the data from these participants were removed from the analyses. In addition, one participant gave identical matches to dress body and lace, neither of which matched the color names they reported for their first view of the dress photo. It was assumed that this participant misunderstood the task and their data were also removed from the analyses. All data for the remaining 32 participants were included in the following analyses (20 females, mean age: 29.3 years, age range: 18.7–60.5 years).

All participants had normal or corrected-to-normal visual acuity and no color vision deficiencies, assessed using Ishihara Color Plates and the Farnsworth-Munsell 100 Hue Test. Each participant received cash compensation for their time.



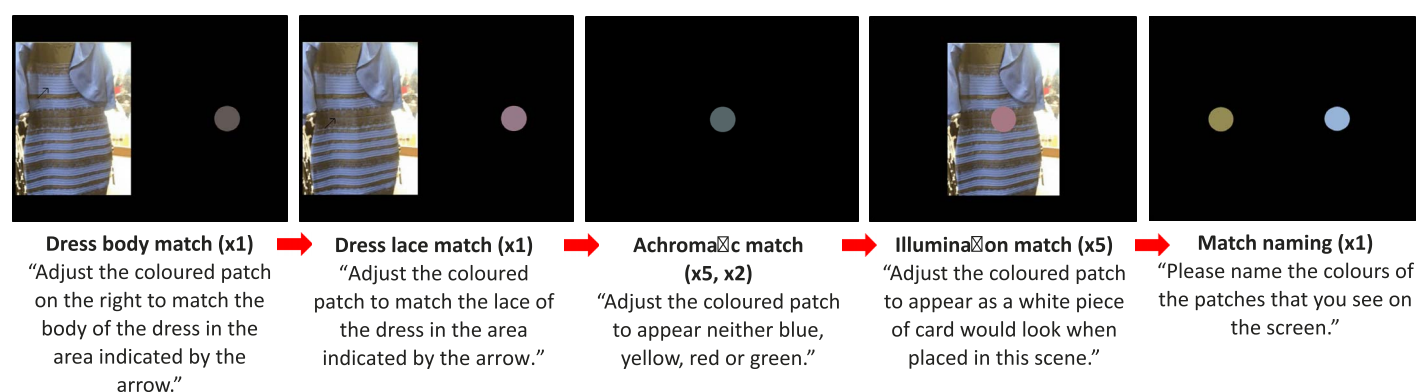


Figure 2. The matching task. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

## Stimuli and apparatus

In the matching task (Figure 2), participants were instructed to match the color appearance of the dress body and lace (one match each), by adjusting the color of a disk presented adjacent to the image of the dress. For each match, a black arrow appeared for 3 s indicating the region of the dress (body or lace) to be matched. Once the arrow disappeared, the matching disk became adjustable. Participants used an Xbox controller to adjust the disk's hue, chroma, and luminance (in steps of 0.1 radians in hue, 1 in chroma and 2  $cd/m^2$  in luminance in HCL color space). For the illumination matches (five matches each), the disk was overlaid on the image and participants were instructed to adjust the disk to appear as if it were a white piece of card present in the scene. Lastly, two conditions of achromatic matches (five of each type) were collected. For both types, a disk was presented in the center of an otherwise black screen and participants adjusted the disk's chromaticity until it appeared neutral—specifically, neither blue, yellow, red, or green. For one condition, the luminance of the disk was fixed at 24  $cd/m^2$ . For the other condition, luminance was fixed at the luminance of that particular participant's dress body match. Hence, for the achromatic settings, only hue and chroma were adjustable by the participant. Dress body and lace matches were always completed first and the order of all remaining matches (illumination and achromatic) was randomized. The chromaticity (and luminance for all but the achromatic matches) was set to a random value at the start of each trial. After all matches were complete, participants were shown two disks on the screen, on the left their dress body match and the right their dress lace match, and asked to name the colors of the disks. The matching disk was the same size on all trials (diameter 4.58 degrees of visual angle). The photograph of the dress was presented in 8-bit color and subsumed 20.41 by 26.99 degrees of visual angle. Stimuli were shown on a 10-bit ASUS Proart LCD

screen (ASUS, Fremont, CA) and matches made with 10-bit resolution. A head rest was used in all tasks to ensure participants maintained a distance of 50 cm from this screen. The monitor was controlled using a 64-bit Windows machine, equipped with an NVIDIA Quadro K600 10-bit graphics card (NVIDIA, Santa Clara, CA), running MATLAB scripts that used Psychtoolbox routines (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). The stimuli were colorimetrically calibrated using a linearized calibration table based on measurements of the monitor primaries made with a Konica Minolta CS2000 spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). Calibration checks were performed regularly throughout the study period and the calibration table was updated when needed. All matches were converted to CIELUV for analysis according to the measured white point of the monitor having coordinates  $(Y, x, y) = (180.23, 0.32, 0.33)$  in CIE  $Yxy$  color space.

The different types of matches specified above are motivated by the following. First, the purpose of the achromatic matches at fixed luminance is to measure any overall bias in the observer's representation of neutral chromaticity at a luminance level that is held constant across observers. Second, collecting matches made at the luminance setting of each individual's dress body match has the purpose of assessing any specific bias due to this particular scene. As the variable that is manipulated here is the fixed luminance setting, it is assumed that bias specific to the scene may be caused by differences in perceived brightness. In both types of adjustments above, the whole scene, and hence the global image statistics, are not available to observers. Under the assumption that computations on global image statistics are necessary to infer the chromaticity and/or brightness of the incident illumination for a particular scene, observers make an illumination match in the context of the scene. In effect, all types of matches can be considered a form of achromatic adjustment. It is expected,

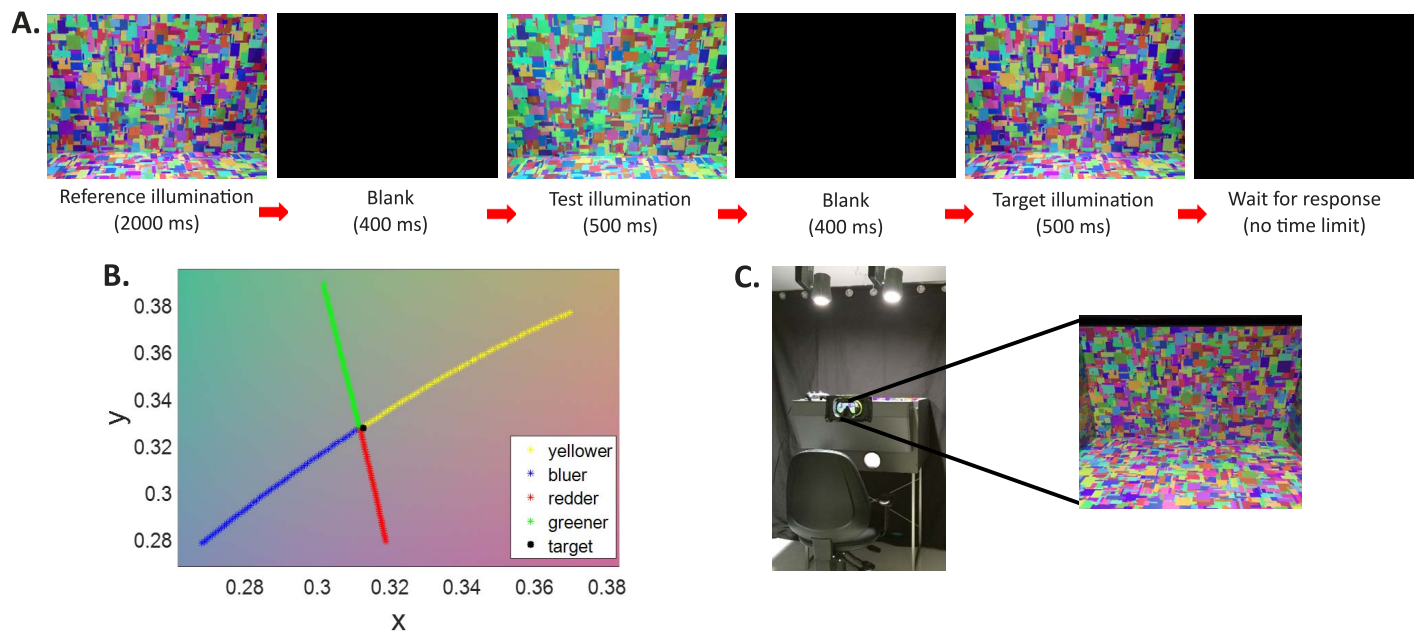


Figure 3. (A) The illumination discrimination task (IDT). (B) The chromaticities of the illuminations used in the experiment plotted in the CIE  $xy$  chromaticity plane. (C) The experimental set up and the participant's view into the Mondrian-lined stimulus box.

however, that the different methods will yield different data sets.

Thresholds for color constancy were obtained using an established illumination discrimination task (IDT; Pearce et al., 2014). The IDT is a two-alternative forced choice task in which participants indicate which of two successively presented illuminations best match a reference (Figure 3A). The reference illumination is a metamer of daylight at 6500K (D65) and one of the comparisons (the target illumination) matches it exactly. The second comparison (the test illumination) varies in chromaticity away from the reference along either the Planckian locus or the line of correlated color temperature (CCT) line to the Planckian locus at D65 (Figure 3B), becoming either bluer, yellower, greener, or redder than the reference. Fifty test illuminations were generated in each direction of change, all at a fixed luminance of  $50 \text{ cd/m}^2$  and parameterized to be approximately one  $\Delta E$  apart in the CIELUV  $u^*v^*$  chromaticity plane. A one-up, three-down, transformed, and weighted staircase procedure (Kaernbach, 1991) is used to find thresholds for correct discrimination of the target and test illuminations along each of the four axes of change. Two staircases are completed for each axis and trials for all eight staircases are interleaved. Thresholds are calculated as the mean of the last two reversals from both staircases for each axis. As a lack of discrimination between the target and test illuminations implies that the surface appearances in the illuminated scene remain stable, higher thresholds are taken to imply improved color constancy for an illumination change

between reference and test relative to lower thresholds. The illuminations were generated by spectrally tuneable 12-channel LED luminaires (produced by the Catalonia Institute for Energy Research, Barcelona, Spain, as prototypes for the EU FP7-funded HI-LED project; [www.hi-led.eu](http://www.hi-led.eu)) and illuminated a Mondrian-lined box (Figure 3C; height 45 cm, width 77.5 cm, depth 64.5 cm), which participants viewed through a viewing port, giving responses using an Xbox controller. The task was run on a 64-bit Windows machine using custom MATLAB scripts. The luminaires' outputs were calibrated by measuring the spectra of each LED primary from a polymer white reflectance tile placed at the back of the viewing box with the Mondrian lining removed, using a Konica Minolta CS2000 spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). Custom MATLAB scripts were used to find sets of weights that define the power of each LED primary needed to produce the smoothest possible illuminations with specified CIE  $Yxy$  tristimulus values (more details on this fitting procedure can be found in Finlayson, Mackiewicz, & Hurlbert, 2014; Pearce et al., 2014; and Radonjic, Pearce et al., 2016).

The MEQ (Horne & Ostberg (1976)) consists of 19 multiple-choice questions. The answers to each question are summed to form a total score, ranging from 16 to 86. Chronotype categories are defined by score subranges as follows: 16–30 “definite evening”; 31–41 “moderate evening”; 42–58 “intermediate”; 59–69 “moderate morning”; and 70–86 “definite morning.”

## Results

### Color names for the dress differ from those assigned to the matched disk colors

The color names that participants reported for the dress body and lace on first view divided into three groups: B/K, W/G, or B/G, in line with previous findings (see Supplementary Material for further details on categorization). Each naming category has similar numbers of participants (B/K = 13, W/G = 11, and B/G = 8). Three observers had never seen the photo of the dress before participating in the experiment.

The color names that participants gave to their disk color matches for the body and lace divided into six groups: the original three groups—B/K, W/G and B/G—and an additional three, purple and gold (P/G), blue and green (B/Gr), and purple and blue (P/B; see Supplementary Material for further details). The number of participants in each group is highly variable. B/G is now the dominant category, with most of the participants who originally named the dress B/G remaining in this group and more than a third of those who originally named the dress B/K and W/G switching to the B/G disk color names group. Thus, participants assign different color names to their matches, when shown them in isolation compared to the color names they use for the dress itself.

In the following analyses (multivariate analyses of variance [MANOVA] and analyses of variance [ANOVA] across disk color name groups), only the four largest naming groups are considered, excluding the two individuals who named the discs B/Gr and B/P, in order to keep the analyses the same as for the dress color names groups (MANOVA requires the number in each group to be at least the number of variables).

### Matches to the dress lace but not dress body differ between dress color naming groups in three-dimensional color space

To determine whether dress body and lace matches vary across the dress-naming groups, we used a three-way MANOVA. This analysis differs from previous tests (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Witzel et al., 2017) in allowing for assessment of differences in a three-dimensional color space (CIE-LUV) by combining the three color coordinates (the three dependent variables) into a composite variable that best represents the differences in the centroids of the three matching groups (B/K, W/G, B/G, the independent variable). For comparison to earlier

studies where differences are considered with respect to each color coordinate separately, we include univariate ANOVA analyses of group differences in the supplementary material.

Color matches to the dress body did not differ significantly across the dress color name groups (MANOVA with dependent variables  $L^*$ ,  $u^*$  and  $v^*$ :  $F(6, 56) = 1.95$ ,  $\Lambda_p = 0.35$ ,  $p = 0.089$ ), and so no composite variable representing the maximal difference between the group centroids was found. Conversely, color matches to the dress lace show a significant multivariate difference across dress color name groups ( $F(6, 56) = 4.03$ ,  $\Lambda_p = 0.60$ ,  $p = 0.002$ ) (Figure 4A through C). The MANOVA analysis yields the composite variable  $d_1 = 0.48L^* + 1.31u^* - 0.43v^*$ . The mean scores for  $d_1$  differ significantly between the B/K and B/G categories ( $p < 0.001$ , Bonferroni corrected) and B/K and W/G categories ( $p = 0.01$ , Bonferroni corrected) (Figure 5A), with lace matches for the B/G and W/G naming groups brighter (increased  $L^*$ ) and more red (higher  $u^*$ ) than those of the B/K group. This is confirmed by pairwise comparisons. The B/K group match the dress lace to be darker (decreased  $L^*$ ), greener (decreased  $u^*$ ), and bluer (decreased  $v^*$ ) than the W/G and B/G groups ( $p < 0.018$  in all cases with a Bonferroni correction). The W/G and B/G dress color name groups did not differ in their matches to the dress lace along any axis of CIELUV color space.

### Disk color names better represent dress body and lace match variability than dress color names

When participants are grouped according to the color names they assign to their disk matches to the dress body and lace, matches differ between groupings more than between the dress color name groups, the former better representing the variation in the matches. With the disk color name grouping employed (Figure 4D through F), there is a significant multivariate difference in the matches to the dress body in CIELUV ( $F(9, 78) = 4.41$ ,  $\Lambda_p = 1.01$ ,  $p < 0.001$ ). Here, the composite variable that best separates the groups is  $d_2 = -0.53L^* - 0.96u^* + 1.82v^*$ , with the P/G group having significantly lower scores on  $d_2$  than the B/G and W/G groups ( $p = 0.005$  and  $p = 0.002$ , Bonferroni corrected) (Figure 5B). Post hoc pairwise comparisons of the  $L^*$  settings of the matches shows that the B/K groups gave a significantly darker dress body match than the W/G and P/G groups (mean differences of 31.30,  $p = 0.007$  and 31.70,  $p = 0.006$ , Bonferroni corrected). The W/G group matched the dress body to significantly higher  $v^*$  values (less blue), than all other groups ( $p < 0.01$ , in all cases with Bonferroni correction).



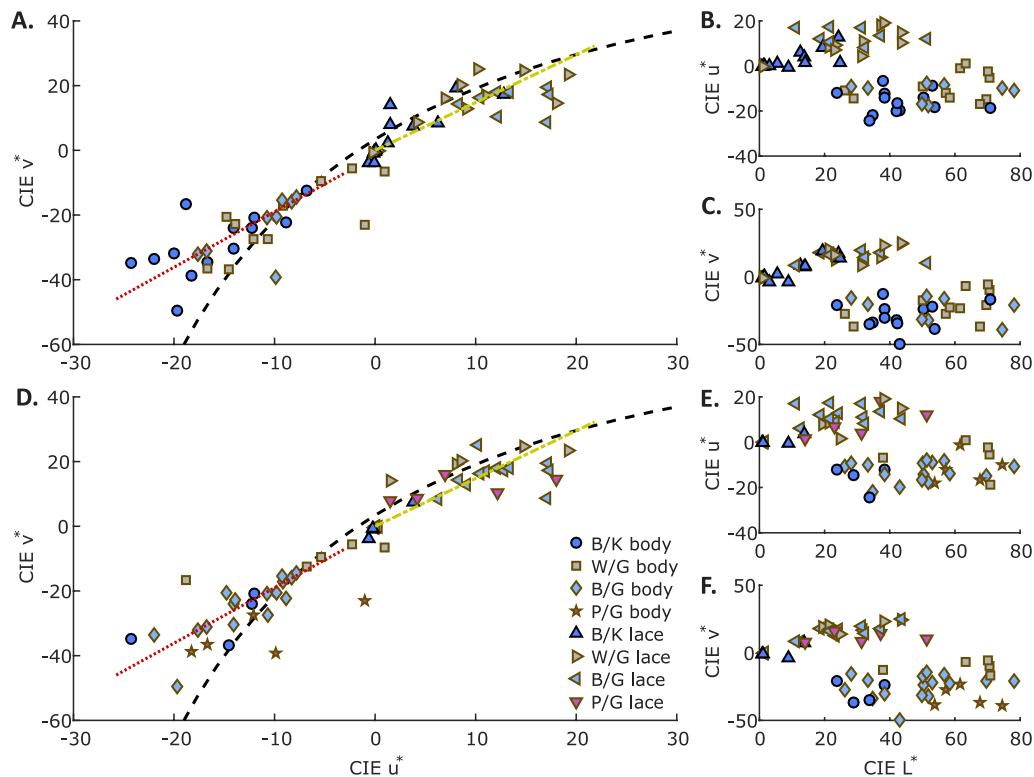


Figure 4. (A–C) The dress body and lace matching data (two points per observer) labeled according to dress color names. Dress body matches are squares, circles, and diamonds. Dress lace matches are differently oriented triangles. (D–F) Same data as in (A–C) labeled according to disk color names, but note that two observers (who named the disks B/Gr and B/P) are omitted from this plot as they were excluded from the corresponding analyses. Black dashed line indicates the Planckian locus. The red dotted line is the first PC of the dress body matches. The yellow dot-dash line is the first PC of the dress lace matches.

Matches to the dress lace also differ significantly across disk color name groups along a multivariate axis defined by the composite variable  $d_3 = -0.68L^* - 0.34u^* + 1.96v^*$  ( $F(9, 78) = 3.26$ ,  $\Lambda_p = 0.82$ ,  $p = 0.002$ ). The B/K group's scores on  $d_3$  indicate that their dress lace matches are darker (decreased  $L^*$ ) and more achromatic (decreased  $v^*$ ) than the matches of the other groups, significantly so in comparison to the W/G group ( $p = 0.002$ , Bonferroni corrected) (Figure 5C). The W/G and P/G group's  $d_3$  scores also differ

significantly ( $p = 0.029$ , Bonferroni corrected). The differences between the B/K and W/G group are further highlighted by the post hoc pairwise comparisons of the  $v^*$  settings, showing that the B/K group gave significantly more achromatic lace matches ( $v^*$  closer to zero) than the W/G group (mean difference of 19.71, Bonferroni corrected). Moreover, the B/K groups  $v^*$  settings were also more achromatic than the B/G group (mean difference of 11.41,  $p = 0.037$ , Bonferroni corrected).

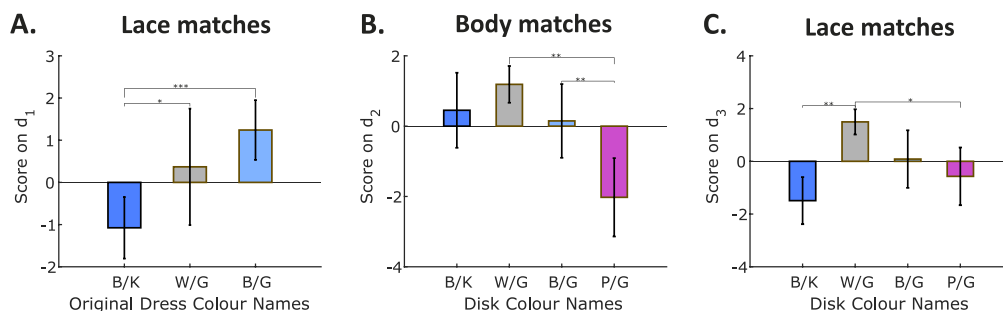


Figure 5. (A) Scores on the composite variable  $d_1$  obtained from the MANOVA analysis on the dress lace color matches categorized by dress color names. (B) Scores on the composite variable  $d_2$  obtained from the MANOVA analysis on the dress body color matches categorized by disk color names. (C) Scores on the composite variable  $d_3$  obtained from the MANOVA analysis on the dress lace color matches categorized by disk color names. Error bars are  $\pm 1SD$ . \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .



## Examining individual variations within and between naming groups: Principal components of the dress body and lace matching data

The matching data (above and in previous studies) show clearly that individuals' color matches to the dress vary along a continuum in color space. Our analyses show that splitting the continuum into groups using the names individuals give to their disk color matches, rather than the color names they give to the dress itself, better represents the variation in the matches. This suggests that the variation in the matches is more suited to being considered on an individual rather than a group level. The MANOVA analyses show that the dress body and lace matches differ across naming groups in a multivariate manner, and the scores on the composite variables retrieved from those analyses might be considered candidates for quantifying the variability. But the composite variables correspond to variation along an axis that optimally represents the differences in the naming category centroids. To capture individual variability, we use principal component (PC) analysis instead to quantify individual variation.

The first PC of the dress body matches is  $PC_b = 0.94L^* + 0.18u^* + 0.3v^*$ , explaining 67.12% of the variation (Figure 4, red dotted line). The first PC of the dress lace matches is  $PC_l = 0.83L^* + 0.31u^* + 0.46v^*$ , explaining 88.92% of the variation (Figure 4, yellow dot-dash line). The two PCs ( $PC_b$  and  $PC_l$ ) are strongly correlated ( $r = 0.783, p < 0.001$ , Figure 6A). The lightness settings of the dress body and lace matches load highly on to the respective components and are themselves strongly correlated ( $r = 0.687, p < 0.001$ , Figure 6B). Thus, lightness seems to drive the association between these matches, with the lightness of one predicting the other. In the following analyses, we characterize how other measures relate to individual variations along these PCs.

## Variation of internal white point explains variation in dress body matches but only when made at the luminance setting of the dress body match

Achromatic settings at the fixed luminance setting of  $24 \text{ cd/m}^2$  (Figure 7A) do not differ between dress or disk color name groups, nor do the achromatic settings produced at the luminance settings of the individual dress body matches (Figure 7D;  $p > 0.332$  in all cases).

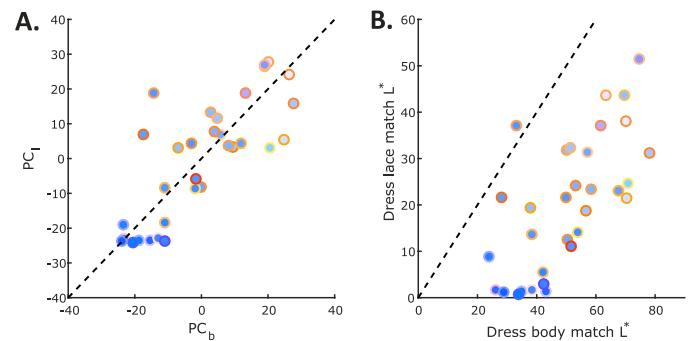


Figure 6. (A) The first PC of the dress lace matches ( $PC_l$ ) plotted against the first PC of the dress body matches ( $PC_b$ ). (B) CIE  $L^*$  setting of the dress lace match plotted against CIE  $L^*$  setting of the dress body match. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE  $Yxy$  values of the matches to sRGB for display.

However, the achromatic settings made at the individual body match luminance levels do explain some of the variation in the dress body and dress lace matches, as revealed by further PC analysis. The first PC of the achromatic settings at fixed luminance of  $24 \text{ cd/m}^2$  is  $PC_f = 0.71u^* + 0.71v^*$ , explaining 67.70% of the variance in achromatic settings; and the first PC of the settings at the dress body match luminance is  $PC_m = 0.42u^* + 0.9v^*$ , explaining 74.63% of the variance.  $PC_m$  in turn explains a small amount of the variation in the dress body and dress lace matches (17.56% and 21.34%, respectively).

While the first PC of the achromatic settings at a fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ) does not correlate with either of the first PCs of the dress body or lace matches ( $PC_b$  and  $PC_l$ ) (Figure 7B through C;  $r = 0.045, p = 0.806$  and  $r = -0.002, p = 0.990$ , respectively), the first PC of the achromatic settings made at the luminance setting of the body match ( $PC_m$ ) does (Figure 7E through F;  $r = -0.419, p = 0.017$  and  $r = -0.462, p = 0.008$ , respectively). This result suggests that the two sets of achromatic matches are unrelated and indeed they are (Figure 8;  $r = 0.180, p = 0.326$ ).

The differences in the two sets of achromatic adjustments may be explained by the luminance settings. For one set of adjustments, all participants were required to make the disk appear achromatic at the same fixed luminance ( $24 \text{ cd/m}^2$ ). For the other, the luminance setting was unique to each individual, and specifically related to that person's perception of the dress. In the latter case, when luminance levels are high, achromatic settings are bluer than when luminance levels are low (negative correlation between  $L^*$  and adjusted  $v^*$  settings,  $r = -0.379, p = 0.032$ ).

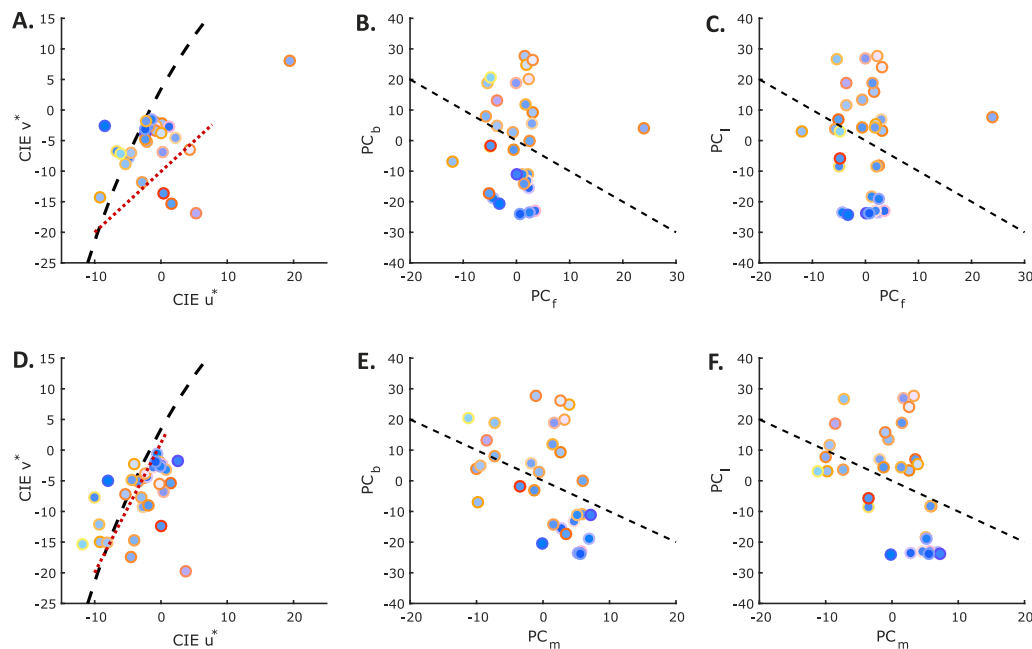


Figure 7. (A) Achromatic settings with luminance fixed at  $24 \text{ cd/m}^2$ . (B) The first PC of the dress body matches ( $PC_b$ ) plotted against the first PC of the achromatic settings at fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ). (C) The first PC of the dress lace matches ( $PC_l$ ) plotted against the first PC of the achromatic settings at fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ). (D) Achromatic settings with luminance set at luminance of dress body match (and therefore varying across participants). (E) The first PC of the dress body matches ( $PC_b$ ) plotted against the first PC of the achromatic settings at the luminance of the dress body matches ( $PC_m$ ). (F) The first PC of the dress lace matches ( $PC_l$ ) plotted against the first PC of the achromatic settings at the luminance of the dress body matches ( $PC_m$ ). Red dotted lines are the first PCs of the data. The black dashed line in (A) and (D) is the Planckian locus. Black dashed lines in remaining plots are lines of perfect negative correlation. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE Yxy values of the matches to sRGB for display.

### Variation in illumination estimates for the dress photo explain some of the variation in dress color matches

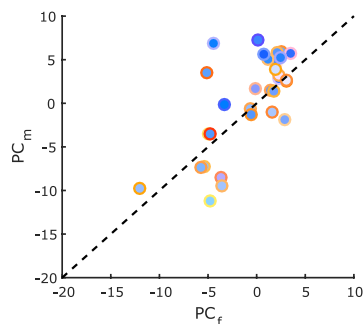


Figure 8. The first PC of the achromatic settings made at the dress body match luminance ( $PC_m$ ) plotted against the first PC of the achromatic settings made at a fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ). The black dashed line is a line of perfect correlation. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE Yxy values of the matches to sRGB for display.

PC analysis of the illumination matches in CIELUV color space uncovers a PC (red dotted line in Figure 9A) that explains 80.21% of the variance defined by  $PC_i = 0.57L^* + 0.357u^* + 0.73v^*$ . While illumination matches do not differ across dress or disk color name groupings ( $p > 0.177$  in all cases), the first PC of the illumination matches ( $PC_i$ ) correlates significantly with the first PC of the dress body matches ( $PC_b$ ; Figure 9B;  $r = -0.395, p = 0.026$ ). There is a similar trend for the first PC of the dress lace matches ( $PC_l$ ; Figure 9C;  $r = -0.343, p = 0.055$ ). These relationships are not predicted by correlations between any of the univariate variables ( $L^*$ ,  $u^*$  or  $v^*$ ), except in the case of the lace matches where the  $v^*$  setting of the dress lace match is negatively correlated with the  $v^*$  setting of the illumination match, with a bluer illumination match implying a yellower lace match. The fact that  $L^*$  and  $v^*$  load highly on to all PCs considered ( $PC_b$ ,  $PC_l$ , and  $PC_i$ ) highlights a relationship between brightness and blue-yellowness perception. Perceiving illuminations as

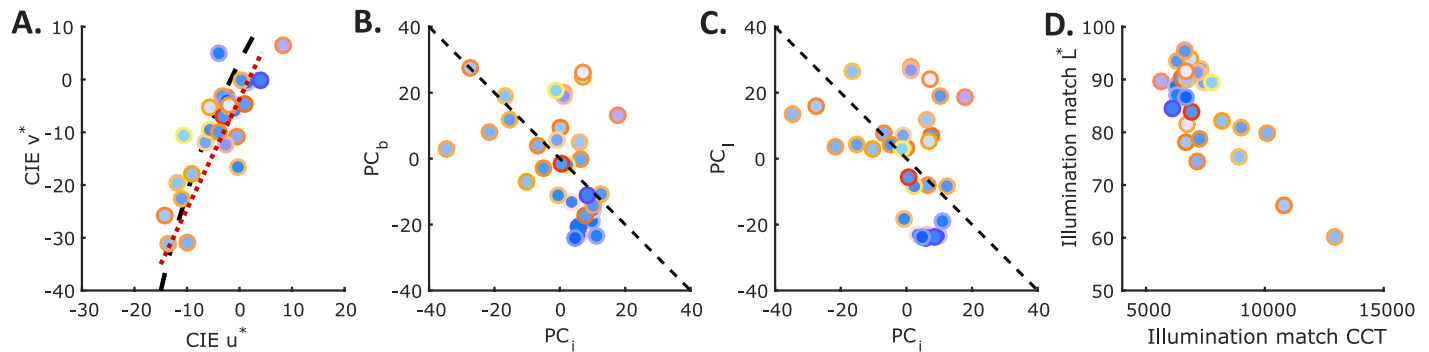


Figure 9. (A) Illumination matches plotted in the CIE  $u^*v^*$  chromaticity plane. The black dashed line is the Planckian locus. The red dotted line is the first PC. (B) The first PC of the dress body matches ( $PC_b$ ) plotted against the first PC of the illumination matches ( $PC_i$ ). (C) The first PC of the dress lace matches ( $PC_l$ ) plotted the first PC of the illumination matches ( $PC_i$ ). (D) Illumination match CIE  $L^*$  setting plotted against illumination match CCT. Black dashed lines in (B–C) are lines of perfect correlation. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE  $Yxy$  values of the matches to sRGB for display.

bluer is linked to perceiving illuminations as darker, and in turn to a brighter, more achromatic dress body match, while yellower is linked to brighter in illumination perception, and to a darker, bluer body match. This relationship between perceived illumination color and brightness is clearly visible in the relationship between CIE  $L^*$  and CCT of the illumination match (Figure 9D,  $r = -0.778$ ,  $p < 0.001$ ). There is a similar relationship between CIE  $L^*$  and CCT of the dress body matches ( $r = -0.502$ ,  $p = 0.003$ ).

The variation in the illumination matches is not related to the variation in the achromatic settings made at a fixed luminance of  $24 \text{ cd/m}^2$  (Figure 10A;  $r = 0.051$ ,  $p = 0.782$ ), but is related to the variation in the achromatic settings made at the luminance setting

of the dress body match (Figure 10B;  $r = 0.367$ ,  $p = 0.039$ ). The PC loadings suggest that lighter and whiter/yellower (higher  $L^*$  and  $v^*$  settings implying a higher  $PC_i$  component score) illumination matches are related to whiter/yellower (higher  $v^*$  settings implying higher  $PC_m$  component score) achromatic settings when the luminance is fixed at the participant's body match luminance setting. This result is difficult to interpret as the relationship is not seen between the underlying values (no relationship between the  $L^*$  or  $v^*$  settings of the two adjustments and no relationship between  $L^*$  settings of the illumination matches and  $v^*$  settings of the achromatic settings,  $|r| < 0.280$ ,  $p > 0.120$ , in all cases). However, there is a relationship between the  $u^*$  settings of the two adjustments ( $r = 0.651$ ,  $p < 0.001$ ).

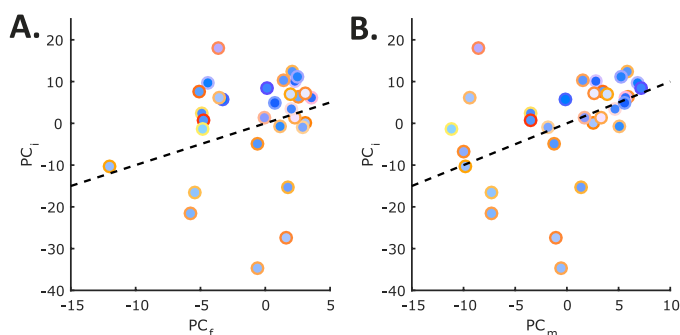


Figure 10. (A) The first PC of the illumination matches ( $PC_i$ ) plotted against the first PC of the achromatic settings at fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ). (B) The first PC of the illumination matches ( $PC_i$ ) plotted against the first PC of the achromatic settings at the dress body match luminance ( $PC_m$ ). Black dashed lines are lines of perfect correlation. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE  $Yxy$  values of the matches to sRGB for display.

### Individual differences in color constancy thresholds assessed via a generic illumination discrimination task do not predict dress color perception

With IDT thresholds grouped according to dress color names (Figure 11A), there is a significant main effect of direction of illumination change on thresholds ( $F(1.97, 57.21) = 17.6$ ,  $p < 0.001$ , with a Greenhouse-Geisser correction). Thresholds for bluer illumination changes are larger than those for yellower, redder, or greener illumination changes ( $p < 0.013$  in all cases with Bonferroni correction) and thresholds for yellower illumination changes are significantly larger than for greener illumination changes ( $p = 0.029$ , with Bonferroni correction). However, there is no interaction effect of group and direction of illumination change on thresholds ( $F(3.95, 57.21) = 0.72$ ,  $p = 0.578$ , with a

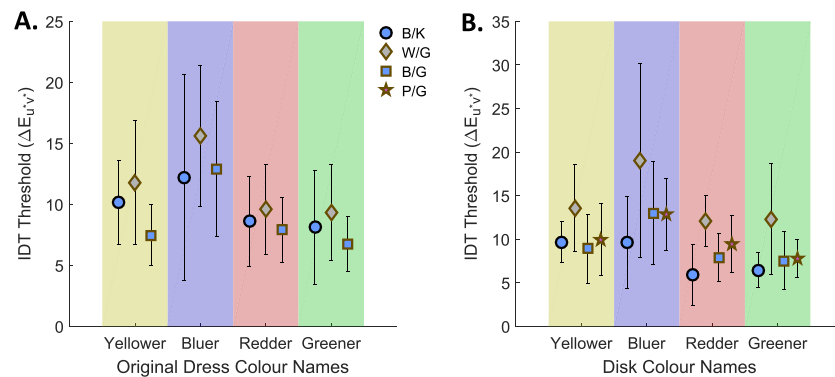


Figure 11. (A) Mean IDT thresholds of the dress color naming groups. (B) Mean IDT thresholds of the disk color naming groups. Coloured underlays indicate the direction of illumination change. Error bars are  $\pm 1$  SD.

Greenhouse-Geisser correction). There is also no main effect of group ( $F(2, 29) = 1.16, p = 0.329$ ).

With participants grouped according to disk color name (Figure 11B), the significant main effect of direction of illumination change remains ( $F(1.97, 51.12) = 10.78, p < 0.001$ , with a Greenhouse-Geisser correction). Thresholds for bluer illumination changes are still larger than for redder and greener illumination changes ( $p < 0.005$  in both cases with Bonferroni correction), but are not larger than those for yellower illumination changes ( $p = 0.172$  with Bonferroni correction). The difference between yellower and greener thresholds still holds ( $p = 0.049$  with Bonferroni correction). Again though, there is no interaction effect ( $F(5.90, 51.12) = 0.48, p = 0.820$ , with a Greenhouse-Geisser correction) or main effect of disk color name ( $F(3, 26) = 2.41, p = 0.090$ ).

There are also no significant correlations between the dress body and lace match PCs ( $PC_b$  and  $PC_i$ ) and IDT thresholds for any direction of illumination change ( $p > 0.1$  in all cases), nor was there a relationship between the PC of the illumination matches ( $PC_i$ ) and any direction of illumination change ( $p > 0.1$  in all cases).

### Does chronotype explain some of the variability in dress color perception?

Scores on the MEQ do not differ significantly across either the dress or disk color name groupings ( $F(2, 29) = 1.14, p = 0.335$  and  $F(3, 26) = 0.73, p = 0.544$ , respectively). In addition, the correlation between MEQ scores and the first PC of the dress body matches was not significant at the 5% level ( $PC_b$ ; Figure 12A;  $r = 0.344, p = 0.054$ ). However, as the effect size is medium (MEQ scores explain 11.83% of the variation in the matching data) and we may have lacked enough power for significance, we consider the implications of such a correlation.  $L^*$  loads

highly onto  $PC_b$  and  $v^*$  loads moderately. This implies that a high score on  $PC_b$  leads to a bright, white/yellow dress body match (i.e., morning types, with high MEQ scores, make brighter, whiter/yellower dress body matches than evening types, with low MEQ score, as we hypothesized). There is also a relationship between MEQ scores and the first PC of the illumination matches ( $PC_i$ ; Figure 12B;  $r = -0.323, p = 0.072$ ), although this correlation is also not significant at the 5% level. Similarly,  $L^*$  and  $v^*$  load highly onto  $PC_i$  implying that a high score on  $PC_i$  indicates a bright, white/yellow illumination match; in other words, evening types make brighter, whiter/yellower illumination matches than morning types.

### Age explains some variability in dress color perception but is also related to MEQ score

While age does not differ significantly across dress color name groups ( $F(2, 29) = 1.12, p = 0.340$ ), there is

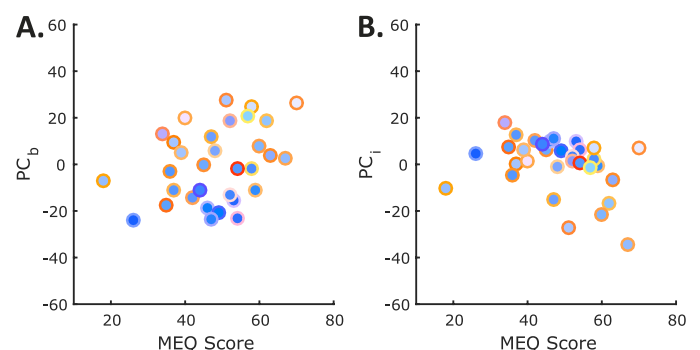


Figure 12. (A) The first PC of the dress body matches ( $PC_b$ ) plotted against MEQ score. (B) The first PC of the illumination matches ( $PC_i$ ) plotted against MEQ score. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE Yxy values of the matches to sRGB for display.



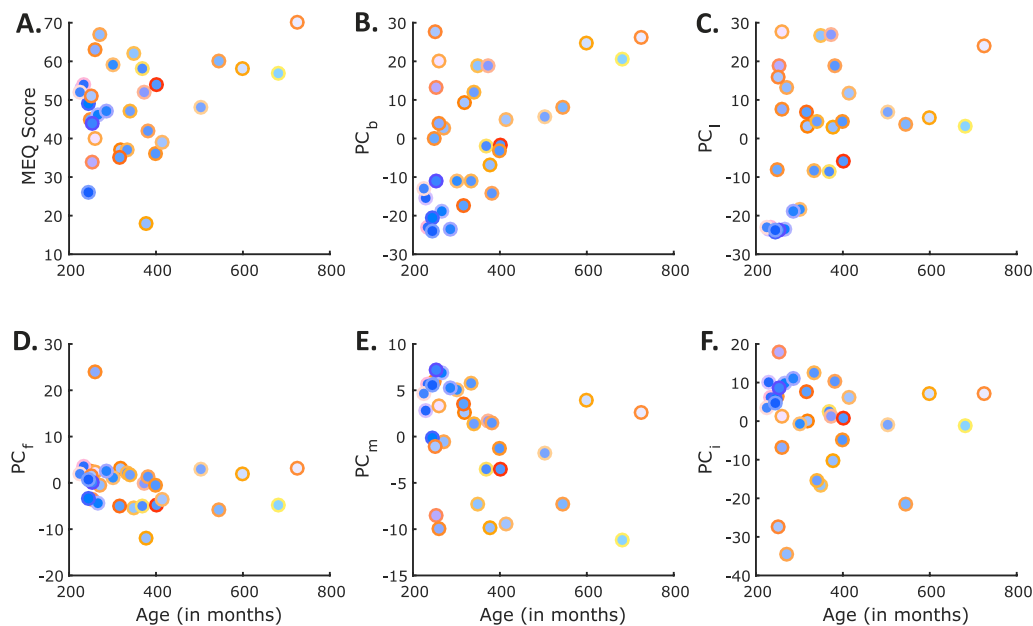


Figure 13. (A) MEQ scores plotted against age. (B) The first PC of the dress body matches ( $PC_b$ ) plotted against age. (C) The first PC of the dress lace matches ( $PC_l$ ) plotted against age. (D) The first PC of the achromatic settings made at a fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ) plotted against age. (E) The first PC of the achromatic settings at the luminance setting of the dress body match ( $PC_m$ ) plotted against age. (F) The first PC of the illumination matches ( $PC_i$ ) plotted against age. All markers are pseudocolored according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the CIE Yxy values of the matches to sRGB for display.

a significant difference in age across disk color name groups ( $F(3, 26) = 5.30, p = 0.005$ ). The W/G disk color name group are significantly older than both the B/K and B/G disk color name groups ( $p = 0.014$  and  $p = 0.006$  with Bonferroni correction, respectively). Moreover, age is related to: MEQ scores (Figure 13A;  $r = 0.316, p = 0.078$ ), with older individuals having higher MEQ scores (more morning type); the first PC of the dress body matches ( $PC_b$ ; Figure 13B;  $r = 0.534, p = 0.002$ ), with older individuals giving brighter (increased  $L^*$ ) and whiter (increased  $v^*$ ) dress body matches (age positively correlates with  $L^*$  and  $v^*$  settings of dress body matches,  $r > 0.427, p < 0.015$  in both cases); the first PC of the dress lace matches ( $PC_l$ ; Figure 13C;  $r = 0.398, p = 0.024$ ), with older individuals giving brighter (increased  $L^*$ ) and yellower (increased  $v^*$ ) dress lace matches (age positively correlates with  $L^*$  and  $v^*$  settings of dress lace matches,  $r > 0.381, p < 0.030$  in both cases); and the first PC of the achromatic settings made at the luminance value of the dress body match ( $PC_m$ ; Figure 13E;  $r = -0.343, p = 0.055$ ). Yet, age is not related to the latter achromatic settings along any single axis of CIELUV color space, nor to the first PC of the achromatic settings made at a fixed luminance of  $24 \text{ cd/m}^2$  ( $PC_f$ ) or of the illumination matches ( $PC_i$ ) (Figure 13D;  $r = -0.191, p = 0.296$  and Figure 13F;  $r = -0.052, p = 0.777$ , respectively).

## Discussion

We set out to test the color constancy explanation of the dress phenomenon. First, we further investigated whether differences in perceived color of the dress are explained by differences in inferred illumination chromaticity, asking whether different prior assumptions were responsible for variations in inferred illumination by using chronotype as a proxy for experience (on the assumption that experience governs the formation of an individual's prior assumptions). Second, we used an established illumination discrimination task to measure color constancy thresholds for illumination changes in our observers, assessing whether individual differences in generic color constancy may explain individual differences in perception. The results of the study support the color constancy hypothesis of the dress phenomenon in demonstrating a relationship between the inferred illumination chromaticity in the image and color matches to the dress (replicating previous results). Correlations, although nonsignificant, between chronotype and dress body matches and between chronotype and inferred illumination chromaticity suggest that unconsciously embedded expectations of illumination characteristics ("illumination priors"), shaped by experience, may act by biasing perception under uncertainty and be responsible for the differences in color names assigned

to the dress. In particular, we find that observers differ not only in the color names they give to the dress, but also in the chromaticity and luminance of the illumination they estimate to be incident on the dress, with the variation in these illumination matches related to both dress body and lace color matches. Further, illumination and dress body matches both show some relationship to MEQ scores, suggesting that chronotype may prove useful as a marker for the chromatic bias of an observer's illumination prior. In addition, our results demonstrate a disconnect between perception and naming. Our observers report different color names for the dress photograph and their isolated color matches, the latter best capturing the variation in the matches.

We find, in agreement with the previous report of Lafer-Sousa et al. (2015), that observers fall into three main groups on the basis of how they named the dress colors on first view: the B/K, W/G, and B/G dress color names groups. We find that color matches to the dress lace differ significantly between the three groups, in both lightness and chromaticity (along both blue/yellow and red/green axes), while color matches to the dress body differ only along a red/green chromatic axis. These results partially replicate the results of two previous studies. First, Lafer-Sousa et al. (2015) reported significant differences in both dress body and lace matches along a lightness and chromatic axis between B/K and W/G naming groups, but did not report any differences between the matches of the B/G group (which they called blue/brown) or their "other" category and any other group. Secondly, Gegenfurtner et al. (2015) reported differences in dress body color matches between B/K and W/G observers but only in luminance settings, a result that has since been replicated by the same group (Toscani, Gegenfurtner, & Doerschner, 2017). The differences in results between studies may be due to differences in sample sizes (53, 15, 38, and 32 for the Lafer-Sousa et al., 2015; Gegenfurtner et al., 2015; Toscani et al., 2017; and current study, respectively), and/or to differences in the luminance and chromaticity of the displayed image and in the chromatic resolution of the matches.

Observers' color matches to the illumination on the dress vary in both chromaticity and luminance and correlate negatively with their color matches to the body of the dress, explaining 15.52% of the variation in the dress body matching data; these results concur with the findings of Witzel, Racey, and O'Regan (2016, 2017) and Toscani et al. (2017) and are related to the findings of Chetverikov and Ivanchei (2016). That is, observers who make illumination matches that are bluer (than the mean) tend to make dress body matches that are less blue, whereas observers who make illumination matches that are less blue tend to make dress body matches that are bluer. A similar trend

exists for the dress lace matching data. The fact that darker illumination matches tend to be more blue than lighter illumination matches fits with the observation that, in nature, indirect lighting and shadows tend to be bluish, due to an absence of direct (yellowish) sunlight (Churma, 1994). Certain visual illusions suggest that the human visual system might have incorporated this relationship, perceiving dimmer areas of images as being indirectly lit or in shadow, and therefore attributing their bluish tints to the illumination rather than the object (Winkler, Spillmann, Werner, & Webster, 2015). It remains an open question whether the bluish tints in brighter areas of the image are also more easily attributed to the illumination than bright yellowish tints. Uchikawa, Morimoto, and Matsumoto (2016) have shown that the optimal color hypothesis (Uchikawa, Fukuda, Kitazawa, & Macleod, 2012) predicts a discrepancy in the color temperature of the inferred illumination in the dress image dependent on the inferred illumination intensity. The peak of the optimal color distribution for the dress image varies with intensity: high intensity implies a more neutral/yellower illumination (light at 5000 K), while low intensity implies a bluer illumination (light at 20,000 K). It remains to be shown though that the optimal color hypothesis is the algorithm adopted by observers to estimate scene illumination.

Achromatic settings made at the luminance of each observer's dress body match explain more of the variation in the dress body and lace matching data than the illumination matches (34.30% and 31.78%, respectively), with the achromatic settings at variable luminance levels showing the opposite trend to the illumination matches: Lighter achromatic settings are more blue than darker achromatic settings, whereas lighter illumination matches are less blue than darker illumination matches. Achromatic settings are commonly used to capture an observer's internal white point, typically assumed to be a measure of the chromaticity of the observer's default neutral illumination (Brainard, 1998). These results suggest that traditional achromatic settings (adjustments of a small patch of fixed luminance to appear achromatic, set against a uniform background) do not reflect the chromaticity of illumination that an observer will estimate for a different and more complex scene; the achromatic settings collected in the present study at a fixed luminance level of  $24 \text{ cd/m}^2$  were not associated to dress color perception or illumination matches. This lack of association agrees with the findings of Witzel et al. (2017), who also found that classical achromatic settings (same fixed luminance across all observers, but with luminance texture), do not predict dress body or lace color matches.

However, the achromatic settings made at the luminance setting of each participant's dress body

match suggest that an observer's internal white point is influenced by luminance, with settings made at higher luminance levels bluer than those at lower luminance levels. It therefore seems that achromatic settings reveal an observer's bias only when an appropriate fixed luminance is chosen for the achromatic adjustment. It is plausible then that when processing a scene radiance image, an observer initially estimates the irradiance level of the illumination, effectively parsing the overall scene radiance (proportional to image irradiance) into material reflectance and illumination irradiance components. Only after this initial parsing does the observer infer the illumination chromaticity, calling on prior knowledge of the relationship between illumination irradiance and chromaticity to do so. The bias lies in the balance of that parsing. Indeed, most of the variation in illumination estimates lie in the luminance settings of the illumination matches.

If we assume that prior knowledge of the relationship between illumination irradiance and chromaticity is a commonality across observers and that different fixed luminance levels are the cause of the variation in  $v^*$  settings seen in the set of achromatic adjustments made at the luminance setting of each observer's dress body match, then we expect a higher fixed luminance to lead to bluer achromatic settings than lower fixed luminance across all observers. In a control experiment (see control experiment in Supplementary Material) a subset of the participants ( $n = 7$ ) made achromatic settings at each of five different luminance levels, spanning the range from minimum to maximum dress body match luminance in the main experiment. There was a clear trend towards increasing blueness (lower  $v^*$ ) with increasing luminance in all participants.

The discrepancy between the blueness-brightness relationship of these variable-luminance achromatic settings and the illumination estimation matches reinforces the conclusion that measurements of subjective white points do not necessarily reflect the chromaticity of the observer's internal default illumination. The illumination matches imply that illuminations that are perceived as brighter are perceived as whiter/yellower, while darker illuminations are perceived as bluer. Similar conclusions from different methods of illumination estimation (Uchikawa et al., 2016; Witzel et al., 2016) support this interpretation. The fact that achromatic settings vary in the opposite way—being bluer at higher luminance levels and yellower at lower luminance levels—suggests that they are measuring something other than the default illumination estimation. Of course, as we pointed out in the Methods section, these differences are to be expected considering the different methods used to obtain the matches.

The observation that achromatic settings vary with luminance has been made before, but the underlying

cause of this variation remains unclear (Chauhan, Perales, Hird, & Wuerger, 2014; Kuriki, 2006, 2015). Conversely, Brainard (1998) found that achromatic settings made at fixed luminance levels lie along a straight line in a three-dimensional cone space, concluding from this that changing luminance did not affect the chromaticity setting of an observer's achromatic point.

The fact that the principal variation in achromatic settings falls roughly on the daylight locus (as also reported by Witzel et al., 2016, 2017; and Witzel, Wuerger, & Hurlbert, 2016), which we find to be more so for the variable-luminance achromatic settings than the  $24\text{cd}/\text{m}^2$ -fixed-luminance settings, may seem to argue that achromatic settings do reflect illumination estimation. Yet if the settings indicate instead the surface chromaticity that appears neutral under natural illumination of that luminance, they would be expected to evince the same variation. The fact that observers require more blue to make the isolated disk appear neutral at higher luminance levels may indicate that they implicitly assign more yellow to the illumination. In the fixed luminance achromatic setting task, the only cue to the irradiance of the illumination—the brightness of the isolated matching disk against a black background—does not vary between observers, and hence there is less cause for variation in the achromatic point between observers. The fact that the variation in achromatic setting at fixed luminance occurs only along an axis roughly orthogonal to the blue-yellow axis also supports the notion that the coupling between assumed irradiance and chromaticity of the illumination falls mainly along the blue-yellow axis. Whatever the correct interpretation of this difference between achromatic settings and illumination estimation, we suggest that care must be taken in future studies to ensure that fixed luminance levels are not influencing behavior in tasks where achromatic settings are taken as a measure of illumination chromaticity. Under the assumption that our illumination matches do capture the variation in illumination estimation across observers, these results suggest that conventional achromatic settings are not indicative of the illumination chromaticity that an observer will estimate on a particular scene.

The main question we address in this study is whether the observer's tendency to infer a particular illumination on the dress is underpinned by a generic bias in illumination priors, as suggested by Lafer-Sousa et al. (2015). Witzel et al. (2017) concluded from similar measurements that the bias is specific to the image, not indicative of a general underlying unconscious expectation of illumination chromaticities. Our additional results suggest that the bias in people who tend to see the dress as lighter and the illumination as darker, is at least partly generic. Scores on the MEQ, a questionnaire that quantifies chronotype, partially predict dress



body and illumination matches. We speculate that the MEQ score provides a quantification of an individual's internal illumination priors as it may predict the illumination chromaticities to which an observer is most frequently exposed due to their interaction with specific environments engendered by their internal body clock. In other words, MEQ scores are indicative of an observer's perceptual biases due to illumination chromaticity estimation. Indeed, we find that observers who score highly on the MEQ and hence are considered more morning-type score higher on the first PC of the dress body match ( $PC_b$ ) than the more evening-type individuals and lower on the first PC of the illumination matches ( $PC_i$ ). This finding supports the hypothesis that an observer's visual system is calibrated for the illuminations in which they find themselves most often, with morning types making bluer illumination matches compared to evening types. It is plausible that morning types are more often exposed to bluer illuminations than evening types as they are more likely to spend time outdoors during the day in bluish daylights (Hernández-Andrés et al., 2001) and less time in yellowish artificial lighting at night (Lafer-Sousa et al., 2015).

In a recent study, Wallisch (2017) also found that morning types (strong larks in that study) are more likely to name the dress W/G than B/K compared to evening types (strong owls). The results are stronger than the relationship we find here. The difference might be due to how Wallisch (2017) parceled observers into chronotype groups. Firstly, the observers are self-categorized and are asked to assign themselves to one of four groups (“strong lark,” “lark,” “strong owl,” and “owl”). In our study we use an established questionnaire (the MEQ) that is widely used to assess chronotype. The MEQ places observers on a scale rather than into distinct groups allowing for more variability that might reduce our ability to show the effect. The same is true for how we represent perception. In our study, observers make color matches to the dress body and lace and the PCs of these matches are correlated with scores on the questionnaire. Again, Wallisch (2017) allowed observers to place themselves into a perceptual category (e.g., W/G observer type).

Age is a confounding factor in our results, with MEQ scores showing a relationship to the PCs of the color matches, but age also differing significantly across the disk color names groups and correlating with the MEQ scores. From this, one may infer that chronotype is not at all related to individual differences in color perception (by being indicative of chromatic bias in illumination priors), but that age is the underlying variable that drives the relationship. The correlation between age and MEQ scores is not a surprise as it is well known that chronotype varies with age (Adan et al., 2012). It would be premature, though, to conclude that age and not experience is the driving factor here as

an individual's illumination priors may change dynamically during their lifetime as their daily experiences also change. On the other hand, aging is known to affect lower level visual factors such as lens optical density (Pokorny, Smith, & Lutze, 1987). To separate these two effects, a study is required that controls for age while sampling from a population of varying chronotypes.

Individual differences in generic color constancy measurements—via the IDT—did not predict individual differences in color matching or naming of the dress. The lack of relationship between dress and illumination matches and IDT thresholds might be because the blue bias is present in all individuals and that the main driver underlying the individual differences in perception of the dress photograph is at a higher level than the illumination discrimination task probes, such as the interpretation of the illumination in the scene. Different types of color constancy tasks (e.g., naming the colors of objects under different illuminations or retrieving the same object under multiple illuminations) may reveal a relationship between color constancy measurements and dress color naming and matching. Alternatively, it might be that the color constancy tasks commonly used in psychophysical experiments do not invoke the use of illumination priors, as the stimuli are well controlled and lack ambiguity. For illumination priors to be used, it seems logical that the incoming sensory information must be somewhat uncertain and that weighting by the prior is necessary to keep perception stable.

Our data also reveal a dissociation between naming and matching, in agreement with our assertion that the distribution of color names assigned to the dress do not fully capture the variety of perceptions experienced by the population. Color appearance matches to the dress body and lace fall along a continuum, not into discrete groupings, confirming previous results (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Witzel et al., 2017). Further, when participants are asked to name their matches presented in isolation, these names differ from those they assign to the dress. The disk color names also predict the dress color matches better than the dress color names, capturing more of the variation in the matching data. The effect is not explained by simple local contrast effects, as in both the naming and matching tasks; the matching disk is surrounded by the same black background, and at the time of making the match, observers seem generally satisfied that the match is representative of how they perceive the colors in the image. A speculative conclusion is that naming the dress B/K, W/G, or B/G involves cognitive as well as phenomenal processes, in that observers do not simply name what they “see,” but that there is a higher order judgment about the color and the linguistic category to which it belongs based on the object with



which it is associated and its surroundings. However, further work is needed to assess this claim.

This observation is relevant to the historical debate over the level on which color constancy exists or may be measured: Is the stability of object color maintained by low-level mechanisms that are inaccessible to conscious influence (Crick & Koch, 1995), or does color constancy require an act of judgment or reasoning at a higher level (Hatfield & Allred, 2012)? The former would entail perfect constancy in the phenomenal sense (i.e., that colors would appear the same under an illumination change and therefore observers would make perfectly equivalent appearance matches, as for example, in the forced-choice matching paradigm of Bramwell & Hurlbert, 1996), whereas the latter would permit the observer to tolerate differences in color appearance while still judging the two surfaces under comparison to be the same (e.g., in the object selection task of Radonjić, Cottaris, & Brainard, 2016). This difference has been brought out empirically in previous color constancy studies, most notably in the “hue-saturation” (appearance) versus “paper” (surface identity) matches of Arend and Reeves (1986) and more recently by Radonjić and Brainard (2016). The matching and naming results reported here suggest that higher level reasoning does play a role in everyday color constancy, and that observers implicitly or explicitly reason about the color of objects in a scene and may report a color name for an object that is not uniquely representative of how it phenomenally appears.

Another possible explanation for the dissociation between the matching and naming results is that the observer is limited in matching the color appearance of a textured object in a complex scene to a uniformly colored matching disk against a black background. Yet here we are asking the observer to make a single match in the same way that the observer gives a single color name to the object. The dress photograph was a social media frenzy because different individuals named the dress different colors—for example, “blue body with black lace” or “white body with gold lace”—using a single color term to describe each part of the dress. This implies that despite a noisy photograph in which dress colors vary on a pixel-by-pixel basis, observers assign a single color to the dress body and lace. What we aim to obtain here through our color appearance matches is exactly that—the color that the dress appears to the observer despite the noise, texture, shading, and other confounds in the image. Obtaining appearance matches in such a way is a standard method in color science and is a method used by other research groups investigating the dress phenomenon (e.g., Gegenfurtner et al., 2015) and other aspects of color perception (e.g., Giesel & Gegenfurtner, 2010; Poirson & Wandell, 1993). Additionally, anecdotal evidence from conversations with

participants after completion of the task suggests they were generally satisfied with their matches and thought their selections accurately represented what they perceived. This reasoning supports the conclusion that the same color appearance elicits different color names depending on its context.

## Conclusions

In summary, these data are supportive of the color constancy explanation of the dress phenomenon, in demonstrating the relationship between inferred illumination and perceived dress color. Furthermore, the results suggest that individual differences in perception of the dress photograph may be partly explained by chromatic bias in illumination priors and that these biases are influenced by factors related to individual experiences. However, we show that a generic measure of individual differences in color constancy does not explain variability in perception of the colors of the dress. In addition, we show that perception and naming may in fact be disconnected. Our results suggest that the color names observers assign to surfaces may depend more on their global perception of the scene rather than only their local surface perception, as observers name their color matches to the dress differently when presented them in isolation from how they name the colors of the dress in the photograph.

*Keywords:* #theDress, color constancy, color vision, individual differences, daylight priors, color naming

## Acknowledgments

This work has been supported by the Wellcome Trust (102562/Z/13/Z to SA). We are grateful to Jay Turner, Naomi Gross, and Daisy Fitzpatrick for assistance with data collection. We also thank several reviewers for their thorough and thoughtful comments.

Commercial relationships: none.

Corresponding author: Stacey Aston.

Email: [stacey.aston1@ncl.ac.uk](mailto:stacey.aston1@ncl.ac.uk).

Address: Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK.

## References

Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., & Randler, C. (2012). Circadian

- typology: A comprehensive review. *Chronobiology International*, 29(9), 1153–1175.
- Allred, S. R. (2012). Approaching color with Bayesian algorithms. In G. Hatfield & S. R. Allred (Eds.), *Visual experience: Sensation, cognition, and constancy* (1st ed.). Cambridge, UK: Oxford University Press.
- Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America, A: Optics and Image Science*, 3(10), 1743–1751.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. Achromatic loci. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, 15(2), 307–325.
- Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, 14(7), 1393–1411.
- Brainard, D. H., & Hurlbert, A. C. (2015). Colour vision: Understanding #theDress. *Current Biology*, 25(13), R551–R554.
- Bramwell, D. I., & Hurlbert, A. C. (1996). Measurements of colour constancy by using a forced-choice matching technique. *Perception*, 25, 229–241.
- Chauhan, T., Perales, E., Hird, E., & Wuerger, S. (2014). The achromatic locus: Effect of navigation direction in color space. *Journal of Vision*, 14(1):25, 1–11, doi:10.1167/14.1.25. [PubMed] [Article]
- Chetverikov, A., & Ivanchei, I. (2016). Seeing “the Dress” in the right light: Perceived colors and inferred light sources. *Perception*, 45(8), 910–930.
- Churma, M. E. (1994). Blue shadows: Physical, physiological, and psychological causes. *Applied Optics*, 33(21), 4719–4722.
- Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375, 121–123.
- Finlayson, G., Mackiewicz, M., & Hurlbert, A. (2014). On calculating metamer sets for spectrally tunable LED illuminators. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, 31(7), 1577–1587.
- Foster, D. H. (2011). Color constancy. *Vision Research*, 51(7), 674–700.
- Gegenfurtner, K. R., Bloj, M., & Toscani, M. (2015). The many colours of “the dress.” *Current Biology*, 25(13), R543–R544.
- Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue, and shape. *Journal of Vision*, 10(9):10, 1–21, doi:10.1167/10.9.10. [PubMed] [Article]
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, 9(11), 1367–8. <http://doi.org/10.1038/nn1794>
- Hatfield, G., & Allred, S. (2012). *Visual experience: Sensation, cognition, and constancy*. Cambridge, UK: Oxford University Press.
- Hernández-Andrés, J., Romero, J., Nieves, J. L., & Lee, R. L. (2001). Color and spectral analysis of daylight in southern Europe. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, 18(6), 1325–1335.
- Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4, 97–110.
- Hurlbert, A. (2007). Quick guide: Colour constancy. *Current Biology*, 17(21), 906–907.
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3), 227–229.
- Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology*, 13, 1–9.
- Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What’s new in Psychtoolbox-3? *Perception*, 36, ECV abstract supplement.
- Kuriki, I. (2006). The loci of achromatic points in a real environment under various illuminant chromaticities. *Vision Research*, 46, 3055–3066.
- Kuriki, I. (2015). Lightness dependence of achromatic loci in color-appearance coordinates. *Frontiers in Psychology*, 6, 67.
- Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by “the dress” photograph. *Current Biology*, 25(13), R545–R546.
- Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS ONE*, 9(2), e87989.
- Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Poirson, A. B., & Wandell, B. A. (1993). Appearance of colored patterns: Pattern-color separability. *Journal of the Optical Society of America*, 10(12), 2458–2470.

- Pokorny, J., Smith, V. C., & Lutze, M. (1987). Aging of the human lens. *Applied Optics*, 26(8), 1437–1440.
- Radonjic, A., & Brainard, D. H. (2016). The nature of instructional effects in color constancy. *American Psychological Association*, 42(2), 847–865.
- Radonjic, A., Cottaris, N. P., & Brainard, D. H. (2016). Color constancy in a naturalistic, goal-directed task. *Journal of Vision*, 15(13):3, 1–21, doi:10.1167/15.13.3. [PubMed] [Article]
- Radonjic, A., Pearce, B., Aston, S., Krieger, A., Cottaris, N. P., Brainard, D. H., & Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, 16(11):2, 1–18, doi:10.1167/16.11.2. [PubMed] [Article]
- Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1458), 1329–1346.
- Sotiropoulos, G., & Seriès, P. (2015). Probabilistic inference and Bayesian priors in visual perception. In *Biologically inspired computer vision* (pp. 201–220). New York: Wiley.
- Toscani, M., Gegenfurtner, K. R., & Doerschner, K. (2017). Differences in illumination estimation in #thedress. *Journal of Vision*, 17(1):22, 1–14, doi:10.1167/17.1.22. [PubMed] [Article]
- Uchikawa, K., Fukuda, K., Kitazawa, Y., & Macleod, D. I. A. (2012). Estimating illuminant color based on luminance balance of surfaces. *Journal of the Optical Society of America*, 29(2), 133–143.
- Uchikawa, K., Morimoto, T., & Matsumoto, T. (2016). Prediction for individual differences in appearance of the dress by the optimal color hypothesis. *Journal of Vision*, 16(12):745, doi:10.1167/16.12.745. [Abstract]
- Wallisch, P. (2017). Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: “The dress.” *Journal of Vision*, 17(4):5, 1–14, doi:10.1167/17.4.5. [PubMed] [Article]
- Webb, A. R. (2006). Considerations for lighting in the built environment: Non-visual effects of light. *Energy and Buildings*, 38(7), 721–727.
- Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue-yellow color perception and in the color of “the dress.” *Current Biology*, 25(13), R547–R548.
- Witzel, C., Racey, C., & O’Regan, J. (2016). Perceived colors of the color-switching dress depend on implicit assumptions about the illumination. *Journal of Vision*, 16(12): 223, doi:10.1167/16.12.223. [Abstract]
- Witzel, C., Racey, C., & O’Regan, J. K. (2017). The most reasonable explanation of “the dress”: Implicit assumptions about illumination. *Journal of Vision*, 17(2):1, 1–19, doi:10.1167/17.2.1. [PubMed] [Article]
- Witzel, C., Wuerger, S., & Hurlbert, A. (2016). Variation of subjective white-points along the daylight axis and the colour of the dress. *Journal of Vision*, 16(12):744, doi:10.1167/16.12.744. [Abstract]
- Wyszecki, G., & Stiles, W. S. (1967). *Color science*. New York: Wiley.